Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing

Year: 2020 | Citations: 916 | Authors: Inioluwa Deborah Raji, A. Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru

Abstract

Rising concern for the societal implications of artificial intelligence systems has inspired a wave of academic and journalistic literature in which deployed systems are audited for harm by investigators from outside the organizations deploying the algorithms. However, it remains challenging for practitioners to identify the harmful repercussions of their own systems prior to deployment, and, once deployed, emergent issues can become difficult or impossible to trace back to their source. In this paper, we introduce a framework for algorithmic auditing that supports artificial intelligence system development end-to-end, to be applied throughout the internal organization development life-cycle. Each stage of the audit yields a set of documents that together form an overall audit report, drawing on an organization's values or principles to assess the fit of decisions made throughout the process. The proposed auditing framework is intended to contribute to closing the accountability gap in the development and deployment of large-scale artificial intelligence systems by embedding a robust process to ensure audit integrity.